# Classification Using SVM and KNN to Predict Voice Pathology

Vikas Mittal
*School of VLSI Design and Embedded Systems*
*NIT, Kurukshetra*
Kurukshetra, India
vikasmittal2k7@rediffmail.com

R.K Sharma
*Department of Electronics and Communication Engineering*
*NIT, Kurukshetra*
Kurukshetra, India
mail2drrks@gmail.com

***Abstract*— The outcome of irregular vocal use is commonly voice pathology. Poor vocal practice, vocal hygiene (because of smoking, dehydration and abusive behavior), and repeated laryngeal infection can lead to worse quality of voice, vocal fatigue, and vocal strain. This research utilizes glottal signal(signal produced by vocal folds) parameters to help out in identify voice disorders linked to vocal folds pathologies. For each recorded speech, the respective glottal signal is acquired from a corpus of male and female speakers of distinct ages using an inverse filtering algorithm. The Mel Frequency Cepstrum Coefficients (MFCCs) also extracted from the voice signal. We select the most relevant as far as pathological / normal discrimination is concerned from the enormous set of parameters obtained. In this paper a new glottal signal parameter (MOQ) is calculated to find Pathological / Normal speech discrimination. Using distinct options, the outcomes are compared: The parameters obtained from the glottal signal, MFCCs and combining both parameters. Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms are used for classifications. Results show that the highest findings of classification, with an average efficiency rise of 3 %, are achieved using the newly studied glottal parameter Maximum Opening Quotient (MOQ), which is a novel outcome and major contribution of this research.**

**Keywords—Pathological voice, Glottal signal parameters, Support vector machine (SVM) and k-Nearest Neighbors (KNN).**

## I. INTRODUCTION

The pathological voice diagnostic science has drawn particular attentions from academic speech processing society over the past centuries. Voice disorders can be divided into three primary classifications: organic, functional and two-category combination. Organic speech disorders split into two groups: structural disorders and neurogenic disorders [1, 2]. Many techniques have been developed in the literature to detect and discriminate pathological voice from normal one [3, 4, 5, and 6].

We are therefore proposing a non-invasive technique to assist clinicians and speech therapists in early identification of vocal fold disease that can enhance evaluation precision. Automatic voice disorder classification is a method based in two main steps. First, the speech utterance extracts a number of parameters. Secondary, the pattern recognition method uses these parameters to classify the voice of the disease and normal one [6, 7]. The glottal signal is used to identify vocal fold-related pathologies [8, 9, and 10]. The glottal signal can be easily obtained inverse filtering of the speech signal [11, 12]. In London and Llorente studies [13], and Pedro et al. [14], discussed the MFCCs were not as efficient in classifying voice pathologies.

Hence, the primary goal of this work is to assess the efficiency of classification models of voice pathologies based on parameters obtained from the glottal signal. In addition, a novel glottal parameter maximum opening quotient (MOQ) is suggested in relation to the maximum opening of the vocal fold, which enables better classification efficiency.

The paper is structures as follows. In section II provides a complete methodology used to classify speech pathologies. Results are provided in section III. Discussions are presented in chapter IV on the outcomes. Section V concludes the paper and provides guidance for future work.

## II. METHODOLOGY

In this paper, the suggested methodology can be summarized in three steps as shown in Figure 1.

(i) Database collection.

(ii) Extraction of MFCCs and glottal parameters.

(iii) Classify the voice / normal pathology

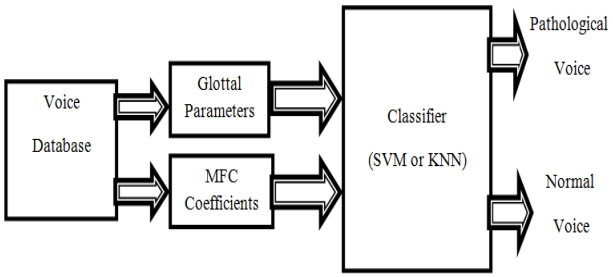


Figure 1. Methodology used to classify pathological voice

### A. Database

The analyzed voices were obtained from 60 speakers divided into 30 dysphonic and 30 normal speakers. To demonstrate the proposed work, we utilized the German database, Saarbrucken Voice Database (SVD) [15, 20] and database developed from MMIMSR, Mullana, and hospital with the help of Dr. Shantanu. The voices recorded in a special sound proof room, using software called “Dr. Speech Software” (Power Sourcing) [16] and an Omni directional microphone. Table 1 show the total number of voice types was 7 (Six pathologies and one normal group).

TABLE 1: PATHOLOGIES FOUND IN THE ANALYZED SAMPLE

| Laryngeal Diseases | No. of cases |
|---|---|
| Normal | 30 |
| Carcinoma | 4 |
| Cyst | 6 |
| Nodule | 8 |
| Unilateral Paralysis | 6 |
| Polyp | 2 |
| Edema | 4 |

### B. The MFFCs

Mel-Frequency Cepstral Coefficients (MFCCs) is a helpful method for extracting characteristics in vibration signals as vibrations contain both linear and non-linear features [17]. MFCC has the following measures for extraction of characteristics: first, a pre-processing of signal is applied to a voice signal. To equalize the exact size, it comprises of a pre-emphasis filter. On each block, a Hamming Window is implemented to reduce the edge effects due to the cutting of windows. A Fast Fourier Transform is applied to the treated signal and smoothed by a sequence of Me1 Scale triangular filters. Then the MFCC is calculated. The MFCCs are obtained and applied to the classifier to find pathological voice in this work for each voice signal. The software PRAAT is used to extract the MFFCs.

### C. Features extracted from the glottal signal

The glottal signal parameters are obtained with the help of a tool box Aparat [19]. The parameters are:

#### (i) Time-domain parameters of glottal signal

The time–domain parameters which can be extracted from the glottal signal are:

(a) Open quotient (OQ): the ratio of the complete moment of the opening vocal folds to the complete moment of the glottal pulse (T) [18].

(b) Closed quotient (CIQ): the ratio of the closing phase parameter to a glottal pulse (T) complete length [18].

c) Amplitude quotient (AQ): The proportion of the glottal signal amplitude (Av) to the glottal signal derivative's minimum value[18].

(d) Normalized amplitude quotient (NAQ): is calculated by the proportion of the AQ to the glottal pulse (T) complete time length [18].

(e) OQ described by the Liljencrants-Fant model (OQa): another opening quotient calculated for inverse filtering by the Liljencrants-Fant model [18].

(f) Quasi open quotient (QOQ): this is the connection between the opening of the glottal signal at the precise oscillation moment and the closing time [18].

(g) Speed quotient (SQ): the ratio between the length of the opening stage and the length of the closing phase [18].

(h) Maximum opening quotient (MOQ): is calculated by the proportion of time interval between the instant when the vocal folds start to oscillate to reach their highest opening point, which is represented by $T_{o1}$ to the total length of glottal cycle or period (T). This parameter indicates that a more asymmetrical glottal flow is produced and time span for abduction is lengthened and adduction of vocal folds is shortened. In case of pathological voice the value of MOQ is increased, which indicates vibration speed of vocal folds slows down. So, in normal (healthy) cases, the value of this parameter will need to be lower. The MOQ is computed as:

$$\text{MOQ}= \frac{T_{o1}}{T} \quad (1)$$

#### (ii) Frequency-domain parameters of glottal signal

(a) Harmonic Difference (DH12): Also known as H1 H2, this is the difference between the first and second harmonic values of the glottal signal [18].

(b) Harmonics richness factor (HRF): refers to the first harmonic (H1) with the energy amount of the other harmonics (Hk) [18]

#### (iii) Parameters that represent differences and disturbances in the fundamental frequency

(a) Jitter: fundamental frequency variations between consecutive cycles of vibration.

(b) Shimmer: glottal flow amplitude variations between consecutive vibrational cycles.

## III. RESULTS

### (i) Analysis of the parameters for classification

The box-plots were facilitated the analysis and better understanding of the variation of glottal parameters, as described in the following subsections and in Figures 2-13.

### (a) Time-domain parameters of the glottal signal

Figures 2-9 demonstrate the respective box-plots obtained from the glottal signal for the so-called time-domain parameters, where some interesting observations can be obtained. OQ1, OQ2, CIQ, AQ and NAQ (Figures 2-6) parameters indicate that when compared to pathologies, normal voices have more intensity and better speech quality. In pathology, the values for the parameters SQ1 and SQ2 are smaller in pathological voices, which indicate a shortening in the structure of vocal folds (Figures 7 and 8).The value of a new investigated parameter MOQ as defined by equation 1 is lower for normal voices, which indicate fast and smooth opening of vocal folds during phonation as compared to one has diseases the vocal folds have slow opening or it acquired more time to reach maximum opening point from the start of oscillation.

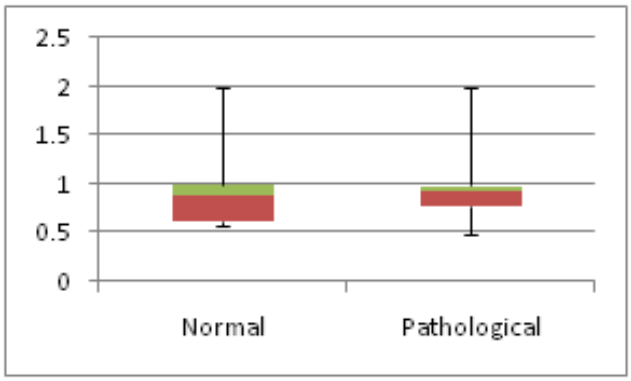


Figure 2. Open Quotient (OQ1)

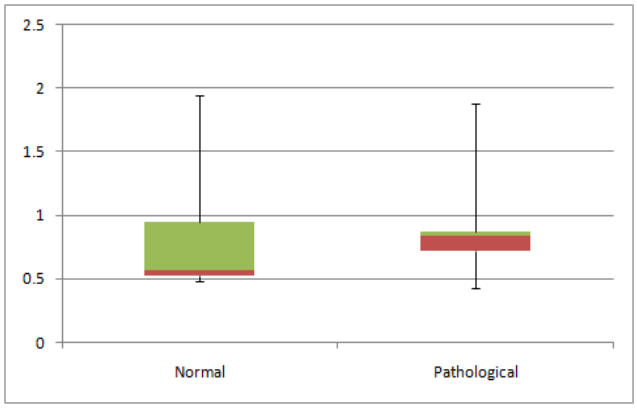


Figure 3. Open Quotient (OQ2)

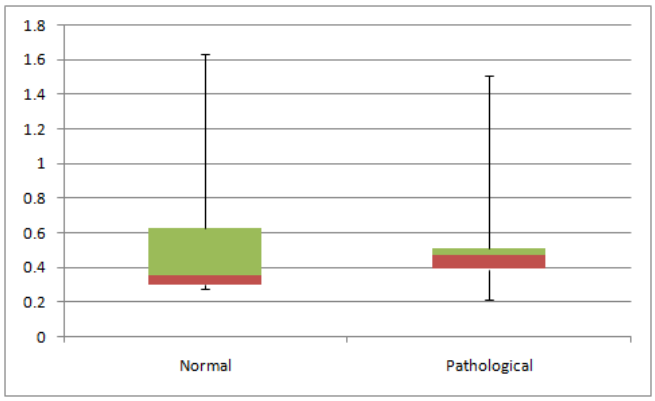


Figure 4. Closed Quotient (CIQ)

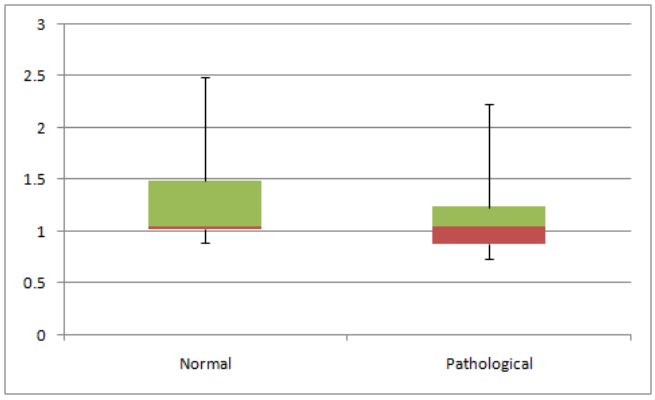


Figure 5. Amplitude Quotient (AQ)

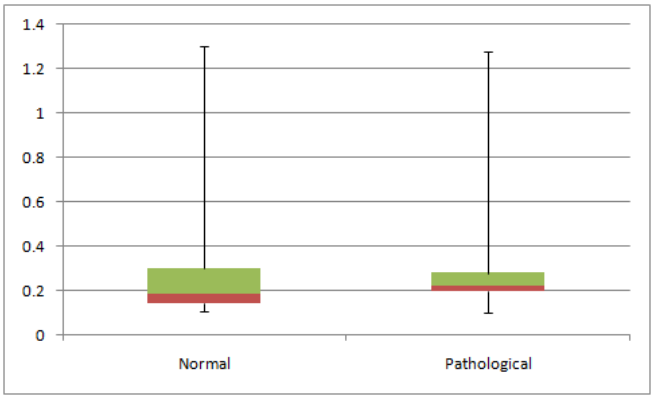


Figure 6. Normalized Amplitude Quotient (NAQ)

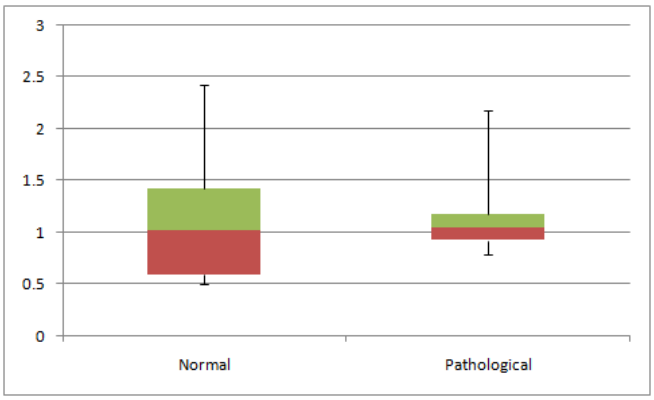


Figure 7. Speed Quotient 1(SQ1)

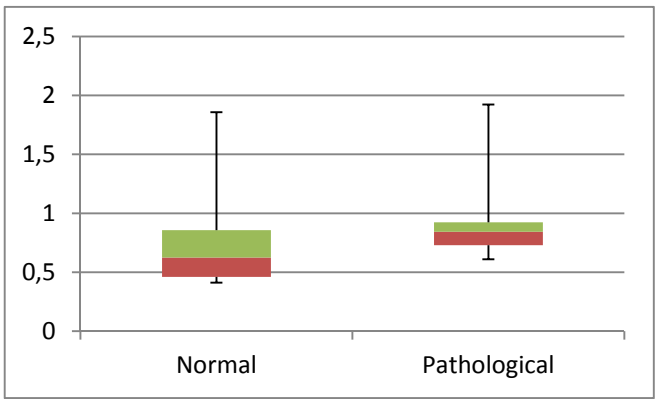


Figure 8. Speed Quotient 2(SQ2)

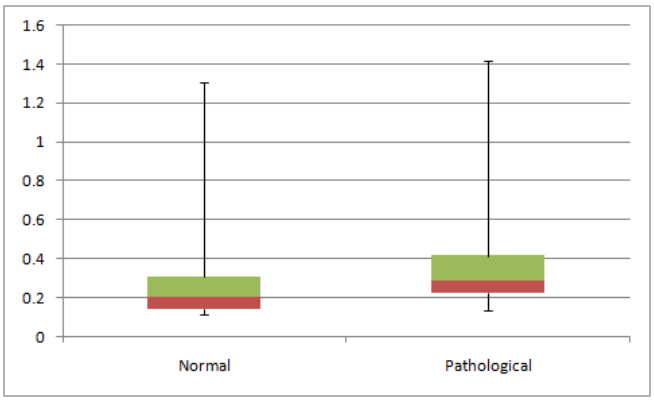


Figure 9. Maximum Opening Quotient (MOQ)

(b) **Frequency-domain parameters**

Figure 10 and 11 shows the resultant box-plots to frequency-domain parameters. The pathological voices have more variations in frequency as compared to normal voices.

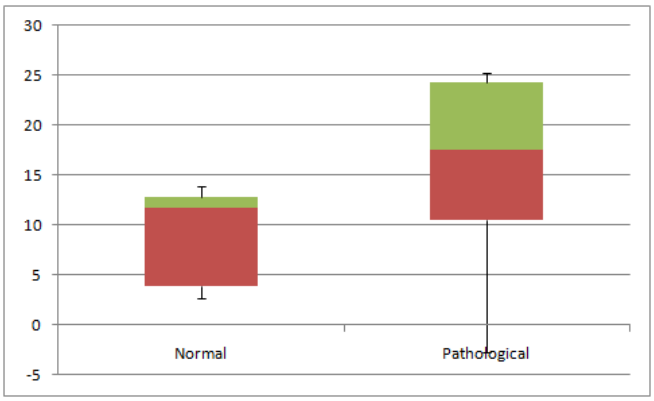


Figure 10. Difference between harmonics (DH12).

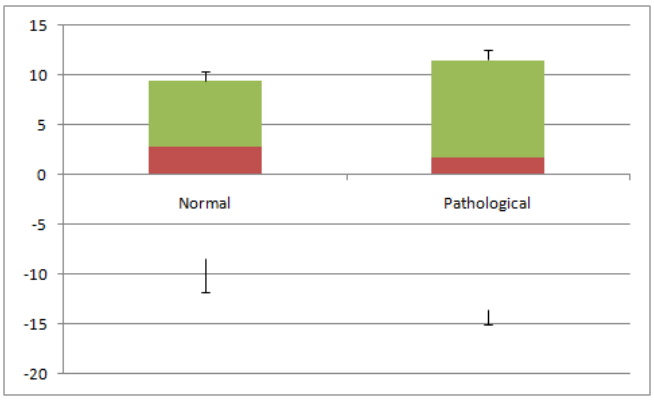


Figure 11. Harmonic richness factor (HRF)

**(c) Parameters related to fundamental frequency**

The shimmer and HNR values in pathological voices are very large, as shown in figures (12-13).

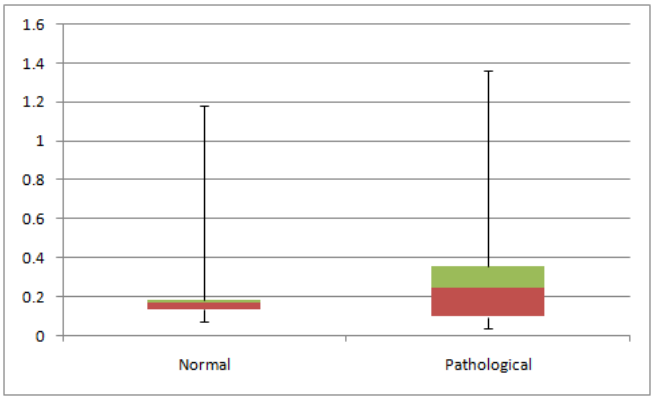


Figure 12. Shimmer

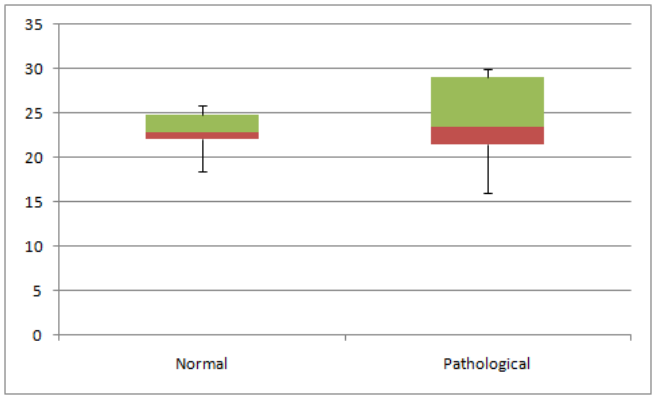


Figure 13. Harmonic to Noise Ratio (HNR)

**(ii) Analysis of Results**

Pathology classification was carried out using various classifiers: SVM and KNN. Six instances were discussed for the input parameters for each classifier: (1) only the MFCCs, (2) only the glottal extracted parameters, (3) Combination of glottal parameters and MFCCs, (4) glottal parameters including new investigated parameter (MOQ), (5) combination of glottal parameters including new investigated parameter (MOQ) and MFCCs, and (6) individual glottal parameters.

(a) Classification Results with the only MFCCs

The table 2 shows that classification accuracy considers SVM as classifier is obtained 71.7% and with KNN classifier is 85%.

Table 2: Confusion matrix of only MFFCs

| SVM Classifier | | | |
|---|---|---|---|
| True Class | Pathological | 25 | 5 |
| | Normal | 12 | 18 |
| | | Pathological | Normal |
| KNN Classifier | | | |
| True Class | Pathological | 23 | 7 |
| | Normal | 2 | 28 |
| | | Pathological | Normal |

(b) Classification Results with only glottal parameters

The table 3 shows that classification accuracy considers SVM as classifier is obtained 86.7% and with KNN classifier is 90%.

Table 3: Confusion matrix of only the glottal parameters

| SVM Classifier | | | |
|---|---|---|---|
| True Class | Pathological | 24 | 6 |
| | Normal | 11 | 19 |
| | | Pathological | Normal |
| KNN Classifier | | | |
| True Class | Pathological | 30 | 0 |
| | Normal | 6 | 24 |
| | | Pathological | Normal |

(c ) Classification Results with combination of the glottal parameters and MFCCs

The input of the classifiers is combination of 16 glottal parameters and 12 MFC coefficients. The confusion matrix is given below in table 4.
The table 4 shows that classification accuracy considers SVM as classifier is obtained 90% and with KNN classifier is 95%.

Table 4: Confusion matrix of MFFCs and glottal parameters

| SVM Classifier | | | |
|---|---|---|---|
| True Class | Pathological | 30 | 0 |
| | Normal | 6 | 24 |
| | | Pathological | Normal |
| KNN Classifier | | | |
| True Class | Pathological | 30 | 0 |
| | Normal | 3 | 27 |
| | | Pathological | Normal |

(d) Classification Results with glottal parameters including new investigated parameter (MOQ)

The input of the classifiers is composed of 16 glottal parameters and plus one new investigated MOQ. The confusion matrix is given below in table 5.
The table 5 shows that classification accuracy considers SVM as classifier is obtained 95% and with KNN classifier is 96.7%.

Table 5: Confusion matrix of glottal parameters and plus one new investigated MOQ

| SVM Classifier | | | |
|---|---|---|---|
| True Class | Pathological | 30 | 0 |
| | Normal | 3 | 27 |
| | | Pathological | Normal |
| KNN Classifier | | | |
| True Class | Pathological | 30 | 0 |
| | Normal | 2 | 28 |
| | | Pathological | Normal |

(e) Classification Results with glottal parameters including new investigated parameter (MOQ) combining with MFCC

The input of the classifiers is collection of glottal parameters including new investigated parameter (MOQ) combining with MFCC. The confusion matrix is given below in table 6.
The table 6 shows that classification accuracy considers SVM as classifier is obtained 93.3% and with KNN classifier is 98.3%.

Table 6: Confusion matrix of glottal parameters and plus one new investigated MOQ with MFCCs

| SVM Classifier | | | |
|---|---|---|---|
| True Class | Pathological | 26 | 4 |
| | Normal | | 30 |
| | | Pathological | Normal |
| KNN Classifier | | | |
| True Class | Pathological | 30 | 0 |
| | Normal | 1 | 28 |
| | | Pathological | Normal |

## IV. Discussion

Table 7 summarizes the outcomes acquired with both the classifiers. The classification was effective, as seen from the outcomes in table 7, concluding that glottal parameters are good discriminators of classifying voice disorders with new investigated glottal parameter (MOQ) and the classification performance improved.

Table 7: Classification of Pathological voice

| Parameters | SVM | KNN |
|---|---|---|
| Only MFFCs | 71.7% | 85% |
| Only glottal parameters | 86.7% | 90% |
| Glottal parameters with MFFCs | 90% | 95% |
| Glottal parameters including MOQ parameter | 95% | 96.7% |
| Glottal parameters with MOQ and MFFCs | 93.3% | 98.3% |

The combination of MFFCs and glottal (with newly investigated glottal parameter MOQ) offered the highest outcomes in classification, with a rise of 3 % in the average performance of various classifiers compared to the outcomes with glottal parameters and MFFCs.

## V. CONCLUSION

A technique has been developed to classify pathological voice using characteristics obtained from glottal and voice signals based on well-known SVM and KNN classifiers.
The main contribution of this work is the use of newly investigated glottal parameter (MOQ). The average accuracy of classification with both classifiers is 95.5 % achieved glottal with MOQ and 91.7% in case of only MOQ parameter, which is better as compared to other cases. The combination of MFFCs and glottal with newly investigated glottal parameter MOQ offered the highest outcomes in classification, with a rise of 3%. Future job includes

preparing databases with audio-visual, audio and video recordings that imply better accuracy of pathological voice.

## ACKNOWLEDGMENT

Authors are thankful to Special Manpower Development Program for Chip to System Design (SMDP-C2SD) initiated by Ministry of Electronics & Information Technology (MeitY), Govt. of India for providing hardware, software and other technical resources in VLSI Design Lab, School of VLSI Design and Embedded Systems, NIT Kurukshetra.

## REFERENCES

[1] R.J. Baken and R.F. Orlikoff, Clinical measurement of speech and voice, Cengage Learning, 2000

[2] P.Schultz, " Vocal fold Cancer," European annals of otorhinolaryngology, head and neck diseases, vol.128. n0.6,pp.301-308, 2011

[3] G.S.Darcio, L.C.Oliveira amd M.Andrea, "Jitter Estimation Algorithms for Detection of Pathological Voices." Euraship Journal of Advances in signal processing, vol. 2009.

[4] T.Dubuisson,T. Dutoit, B.Gosselin, and M.Remacle,"On the use of the correlation between acoustic descriptors for the normal/pathological discrimination,"EURASIP Journal on advances in signal processing, vol.2009.

[5] V.Sellam and J.Jagadeesan, " Classification of on Processing, Normal and Pathological Voice Using SVM and RBFNN," Journal of signal and information Processsing, vol.5,pp.1-7,2014.

[6] M. Vasilakis and Y. Stylianou, "Voice Pathology detection based on short-term jitter estimation in running speech." Folia Phoniatrica and Logopaedica, vol.61, no.3,pp.153-170. 2009.

[7] J. Wang and C.Jo, "Performance of gaussian mixture models as a classifier for pathological voice," in Proc. International Conference on Speech Science and Technology, 2006.

[8] J. Cheolwoo,"Source Analysis of Pathological Voice," in Proc. International multiconference of engineers and computer scientists,2010.

[9] Hariharan M.Paulraj MP, Yaacob S. identifiaction of Vocal Fold Pathology Based on Mel Frequrncy Band Energy Coefficient and singular Value Decomposition. Signal and Image Prcessing Applications(ICSIPA);2009:514-517.

[10] Kohler M,Mendoza LF,Lazo J,Vellasco M,Cataldo E. Classification of voice pathologies using glottal signal parameters. Proc. Of Brazilian Congress on Computational Intelleligence(CBIC);2011.

[11] Verdonck-de Leeuw IM, Mahieu HF. Vocal aging and the impact on daily life: a longitudinal study.J Voice.2004;18:193-202.

[12] Steffen N,Pedrosa VV,Kazuo R,Pontes P. Modifications of veestitublar fol shape from respiration to phonation in unilateral vocal fold shape from respiration to phonation in unilaeral vocal fold paralysis. J Voice.2009;25:111-113.

[13] Londono J, Liorente J. An improved method of voice pathology detection by means of a HMM-based feature space transformation.Pattern Recogn. 2010;43:3100-3112.

[14] Pedro Gomez-Vilda et al. Glottal source biometrical signature for voice pathology detection. Speech Communication 2009;51:759-781.

[15] W.J.Barry and M.Putzer, " Saarbrucken Voice Database," Institute of Pjonetics,Univ. of Saarland,http://www.stimmdatenbank.coli.uni-saarland.de/.

[16] Dr. Speech Software, Tiger DRS,China. Available at: http://www.drspeech.com/, Accessed June 15,2015.

[17] Nelwamondo, Fulufhelo V., and Tshilidzi Marwala. "Faults Detection Using Gaussian Mixture Models, Mel-Frequency Cepstral Coefflcients and Kurtosis." *2006 IEEE International Conference on Systems, MAan, and Cybernetics October 8-11, 2006, Taipei, Taiwan*. 1-4244-0100-3/06. (2006): 290-295. Print.

[18] Pulakka H. Analysis of Human Voice Production Using Inverse Filtering , High-Speed Imaging, and Electroglottography, MSc Dissertation,University of Technology Helsinki;2005.

[19] Software Aparat, Available at: http://aparat.sourceforge.net/index.php/main_page, Helsinki University of Technolgoly Laboratory of Acoustics and Audio Signal Processing, Accesessed June 15, 2015.

[20] Vikas Mittal, R.K.Sharma. Detection and Classification of voice pathology using electrical parameters,International Journal of Engineeering and Advanced Technology(IJEAT) 2019;8:3836-39